

# CONSTRUCTS FOR HOMOGENOUSLY PROCESSED PREPARATIONS OF BETA SITE APP-CLEAVING ENZYME

## PRIORITY CLAIM

[0001] The present application claims priority under 35 U.S.C. § 119(e) to U.S.S.N. 60/430,984, filed December 4, 2002; the entire contents of which are incorporated herein by reference.

## FIELD OF THE INVENTION

[0002] This invention provides constructs and polypeptides for producing preparations of homogeneously processed BACE.

## BACKGROUND

[0003]  $\beta$ -secretases play an important role in disease. These proteins cleave the amyloid precursor protein (APP) at its  $\beta$ -secretase site; APP is subsequently cleaved at its  $\gamma$ -secretase site by  $\gamma$ -secretases, leading to the liberation of the A $\beta$  peptide. Cerebral deposition of the A $\beta$  peptide is part of the pathology of Alzheimer's disease and is also frequently observed in Down syndrome. Given the significance of  $\beta$ -secretases as disease targets, structural information obtained about their interaction with inhibitors is valuable in the design of compounds that inhibit  $\beta$ -secretase activity. An example of a  $\beta$ -secretase is beta site APP-cleaving enzyme1, BACE1.

[0004] BACE1 is a transmembrane protein that operates in the lumen of the Golgi apparatus [Vassar R., et al., *Science* **286**:735-741(1999)]. The full-length protein, also known as the preprosequence, comprises a presequence, a prodomain, a linker region, a protease domain, a transmembrane domain, and a cytosolic domain. The preproprotein sequence of the longest isoform of BACE1, isoform A, is shown in SEQ ID NO:1. There are three other isoforms of BACE1- isoform B, isoform C, and isoform D- that lack residues 190-214, 146-189, and 146-214, respectively, relative to SEQ ID NO:1.

[0005] SEQ ID NO:1

```
1 MAQALPWLLL WMGAGVLP AH GTQHGIRLPL RSLGGAPLG LRLPRETDEE PEEPGRRGSF
61 VEMVDNLRGK SGQGYVEMT VGSPPTLNI LVDTGSSNFA VGAAPHPFLH RYYQRQLSST
121 YRDLRKGVYV PYTGKWEGE LGTDLVSIPH GPNVTVRANI AAITESDKFF INGSNWEGIL
181 GLAYAEIARP DDSLEPPFDS LVKQTHV PNL FSLQLCGAGF PLNQSEVLAS VGGSMIIGGI
241 DHSLYTGSLW YTPIRREWY EVIIVRVEIN GQDLKMDCKE YNYDKSIVDS GTTNLRLPKK
```

301 VFEEAVKSIK AASSTEKFPD GFWLGEQLVC WQAGTTPWNI FPMVISLYLMG EVTNQSFRT  
 361 ILPQQYLRPV EDVATSQDDC YKFAISQSST GTVMGAVIME GFYVVFDRAR KRIGFAVSAC  
 421 HVHDEFRTAA VEGPFVTLDL EDCGYNIPQT DESTLMTIAY VMAAICALFM LPLCLMVCQW  
 481 RCLRCLRQQH DDFADDISLL K

[0006] To date, there has been some degree of variability in the stated boundaries of some of the domains of BACE1; the amino acid numbering in the following is for isoform A. There appears to be consensus on the definitions of the presequence (1-21), the prodomain (22-45), and mature BACE1 (46-501). Boundaries of the transmembrane domain and the cytosolic domain have been within one to three amino acids of each other. For example, both SBASE and Swiss Prot have the potential transmembrane domain and the potential cytosolic domain consisting of residues 458-478 and 479-501, respectively. The boundaries for these two domains have also been described as 461-477, and 478-501, respectively [Shi, X.-P., et al., *J Biol Chem* **276**:10366-10373 (2001)]. The protease domain has shown the most variability in its definition. It has alternately been placed between residues 46-460 [Shi, X.-P., et al., *J Biol Chem* **276**:10366-10373 (2001)], 73-390 (S-BASE), and 74-416 ([NP\\_036236](#)). Additionally, a human BACE1 enzyme construct having an estimated protease domain ending at residue 419 showed activity against substrates [Lin, X., et al., *Proc. Natl. Acad. Sci. USA* **97**:1456-1460 (2000)].

[0007] Other features of BACE1, isoform A, are catalytic Asp residues, D93 and D289, which are each part of a D(S/T)G motif; N-linked glycosylation sites at N153, N172, N223, and N354; and disulfide bonds formed between C216 and C420, between C278 and C443, and between C330 and C380 [Charlwood, J., et al., *J Biol Chem* **276**:16739-16748 (2001)]. Additionally, mature BACE1 is phosphorylated at S498 in the cytosolic domain [Walter, J., et al., *J Biol Chem* **276**:14634-14641 (2001)] and palmitoylated at residues C478, C482 and C485 in the cytosolic domain [Benjannet, S., et al., *J. Biol. Chem.* **276**:10879-10887(2001)].

[0008] Soluble proteins containing the pre, pro, and protease domains can be expressed in bacterial and eukaryotic cell types [Vassar, R., et al., *Science* **286**:735-741(1999); Lin, X., et al., *Proc. Natl. Acad. Sci. USA* **97**:1456-1460 (2000); Mallender, W. D., et al., *Mol Pharmacol.* **59**:619-626 (2001)]. The prodomain is required for folding in secreted expression systems [Shi, X.-P., et al., *J Biol Chem* **276**:10366-10373 (2001)] and for refolding of protein expressed as inclusion bodies in *E. coli*. BACE expressed in Chinese hamster ovary cells is processed to yield an N-terminus at amino acid residue 46 (from the initiator Met residue), by a furin-like proprotein convertase (PC), which is believed to be responsible for the normal physiological processing *in vivo* [Benjannet, S., et al., *J. Biol. Chem.* **276**:10879-10887 (2001); Creemers, J.

W., et al., *J. Biol. Chem.* **276**:4211-4217 (2001)]. Additionally, the substrate sequence preferences of purified BACE1 have been examined [See, for example, Turner, R. T. 3<sup>rd</sup>, et al., *Biochemistry* **40**:10001-10006 (2001); Gruninger-Leitch, F., et al., *J. Biol. Chem.* **277**:4687-4693 (2002)].

[0009] Proteins produced by *E. coli* expression are generally preferred for X-ray crystallographic structure determination over those produced from insect or mammalian cells, because the former are devoid of heterogeneous glycosylation. However, previous attempts to express proBACE in *E. coli* and refold into active enzyme did not, however, generate crystals as a free enzyme or in complex with several small-molecule inhibitors. Hong, L., et al. [*Science* **290**:150-153 (2000)] have determined the crystal structure of soluble BACE1 produced from *E. coli*, using protein that had been processed heterogeneously after amino acids Gly40 and Arg42 -numbered relative to SEQ ID NO:1-presumably by a contaminating *E. coli* protease during purification. This processing event was not reproducible and would not be expected to be consistent from preparation to preparation. It is therefore useful to engineer proBACE constructs that can be processed homogeneously and robustly.

#### SUMMARY OF THE INVENTION

[0010] This invention is directed to engineered polypeptides having BACE activity. The polypeptides are properly folded and are cleaved at an engineered cleavage site *in vitro*, producing homogeneous preparations of purified polypeptides having BACE activity.

[0011] The invention further pertains to nucleic acids, expression vectors, and host cells comprising the expression vectors for making the polypeptides having BACE activity.

#### BRIEF DESCRIPTION OF THE DRAWING

[0012] Figure 1 shows a deconvoluted mass spectrum of the processed product from the EINL-BACE1 polypeptide of SEQ ID NO:84. The apparent mass of the processed product is 45,558.0 Da, which is within 3.7 Da of the predicted mass of 45,561.7 Da for residues 25-433 of SEQ ID NO:84, corresponding to cleavage immediately C-terminal to the Leu24 of SEQ ID NO:84. Leu24 of SEQ ID NO:84 corresponds to Arg45 of the preprosequence (SEQ ID NO:1) of human BACE1.

## DESCRIPTION OF CERTAIN PREFERRED EMBODIMENTS OF THE INVENTION

### [0013] *Definitions*

[0014] “Acidic pH” as used herein refers to a pH less than 6, more preferably less than about 5.5, and most preferably between 4 and 5.5.

[0015] “BACE activity” is defined as the ability of a polypeptide under the *in vitro* conditions described by Mallender, W. D., et al., [*Mol Pharmacol.* 59:619-626 (2001)] to cleave the 29 amino acid peptide described therein comprising the Swedish mutant APP sequence SEVNLDAEFR (SEQ ID NO:2) at the scissile bond located between the Leu and the Asp contained within SEQ ID NO:2 with a specific activity of at least about 20 nmol/min/mg, preferably about 90 nmol/min/mg, more preferably about 200 nmol/min/mg, and even more preferably about 400 nmol/min/mg, and most preferably about 900 nmol/min/mg or higher.

[0016] The “prodomain” comprises at least six contiguous amino acids of SEQ ID NO:3. SEQ ID NO:3 corresponds to residues 22-37 of SEQ ID NO:1, or residues 22-37 of human BACE1.

TQHGIRLPLRSGLGGA

SEQ ID NO:3

[0017] In one embodiment, the prodomain comprises at least seven contiguous amino acids of SEQ ID NO:3. In another embodiment, the prodomain comprises at least eight contiguous amino acids of SEQ ID NO:3. In yet another embodiment, the prodomain comprises at least ten contiguous amino acids of SEQ ID NO:3. In still another embodiment, the prodomain comprises at least twelve contiguous amino acids of SEQ ID NO:3. In another embodiment, the prodomain comprises the full length of SEQ ID NO:3. In another embodiment, the prodomain comprises residues 22-41 of SEQ ID NO:1. In another embodiment, the prodomain comprises residues 22-45 of SEQ ID NO:1.

[0018] An “engineered cleavage site” as used herein is an autoproteolysis site or an exogenous protease cleavage site that has been introduced into a polypeptide having BACE activity.

[0019] An “autoproteolysis site” as used herein is defined as comprising the sequence X1-X2-X3-X4-X5,

wherein

X1 is Glu or Gln;

X2 is Leu, Ile or Val,

X3 is Asn, Asp or Met,

X4 is Leu or Phe, and

X5 is Glu, Met, Gln, Ser, Ala or Asp.

[0020] An “exogenous protease” as used herein is a polypeptide having activity other than BACE activity.

[0021] An “exogenous protease cleavage site” as used herein is a sequence comprising at least two amino acids that is cleaved by an exogenous protease. Examples of exogenous protease cleavage sites are a thrombin cleavage site, a tobacco etch virus cleavage site, a Genenase I cleavage site, an Enterokinase cleavage site, a Granzyme B cleavage site, a turnip mosaic virus protease NIa cleavage site, and a Factor Xa cleavage site.

[0022] A “thrombin cleavage site” as used herein is defined as the sequence PR, or as the sequence GRG, where in either case the scissile bond is located immediately after the Arg. More preferably it is the sequence LVPRGS (SEQ ID NO:4), where the scissile bond is located between the Arg and the Gly.

[0023] A “tobacco etch virus protease cleavage site”, used interchangeably herein with a “TEV protease cleavage site”, are each defined as the sequence ENLYFNX (SEQ ID NO:5) wherein the scissile bond is located immediately after the second Asn. More preferably, it is a sequence selected from the group consisting of ENLYFNG (SEQ ID NO:6) and ENLYFNA (SEQ ID NO:7), wherein for each case, the scissile bond is located immediately after the second Asn.

[0024] A “Genenase I cleavage site” as used herein is defined as a sequence selected from the group consisting of AXHYA (SEQ ID NO:8); AXHFA (SEQ ID NO:9); AXHLA (SEQ ID NO:10); FXHYA (SEQ ID NO:11); FXHFA (SEQ ID NO:12); FXHLA (SEQ ID NO:13); LXHYA (SEQ ID NO:14); LXHFA (SEQ ID NO:15); LXHLA (SEQ ID NO:16); YXHYA (SEQ ID NO:17); YXHFA (SEQ ID NO:18); and YXHLA (SEQ ID NO:19), wherein for each case, the scissile bond is located immediately before the last Ala of the sequence. More preferably it is the sequence PGAAHYA (SEQ ID NO:20) wherein the scissile bond is located immediately before the last Ala of the sequence.

[0025] An “Enterokinase cleavage site” as used herein refers to a sequence selected from the group consisting of DDK, DEK, EDK, and EEK, wherein for each case, the scissile bond is located immediately after the Lys; or a sequence selected from the group consisting of DDR, DER, EDR, and EER, wherein for each case, the scissile bond is located immediately after the Arg; and wherein the aforementioned sequences are not immediately followed by a Pro. More

preferably, the enterokinase cleavage site is a sequence selected from the group consisting of DDDDK (SEQ ID NO:21); DEDDK (SEQ ID NO:22); DDEDK (SEQ ID NO:23); DDDEK (SEQ ID NO:24); EDDDK (SEQ ID NO:25); DEEDK (SEQ ID NO:26); DEDEK (SEQ ID NO:27); DDEEK (SEQ ID NO:28); EDDEK (SEQ ID NO:29); EEDDK (SEQ ID NO:30); EDEDK (SEQ ID NO:31); EEEDK (SEQ ID NO:32); EEDEK (SEQ ID NO:33); EDEEK (SEQ ID NO:34); DEEEK (SEQ ID NO:35); and EEEEEK (SEQ ID NO:36); wherein for each case, the scissile bond is located immediately after the Lys, and wherein the aforementioned sequences are not immediately followed by a Pro.

[0026] A “Granzyme B cleavage site” as used herein refers to a sequence selected from the group consisting of IEXD, IMXD, IQXD, VEXD, VMXD, and VQXD, wherein for each case the scissile bond is located immediately after the Asp in the last position of the sequence. Preferably, the Granzyme B cleavage site is a sequence selected from the group consisting of IEXDXG (SEQ ID NO:37); IEXDXA (SEQ ID NO:38); IMXDXG (SEQ ID NO:39); IMXDXA (SEQ ID NO:40); IQXDXG (SEQ ID NO:41); IQXDXA (SEQ ID NO:42); VEXDXG (SEQ ID NO:43); VEXDXA (SEQ ID NO:44); VMXDXG (SEQ ID NO:45); VMXDXA (SEQ ID NO:46); VQXDXG (SEQ ID NO:47); and VQXDXA (SEQ ID NO:48), wherein for each case the scissile bond is located immediately after the Asp in the fourth position of the sequence.

[0027] A “TuMV NIa protease cleavage site” used interchangeably herein with “turnip mosaic virus protease NIa cleavage site” is defined as the sequence VXHQ, wherein the scissile bond is located immediately after the Gln. More preferably, the TuMV NIa protease cleavage site is the sequence VRHQS (SEQ ID NO:49), wherein the scissile bond is located immediately after the Gln.

[0028] A “Factor Xa cleavage site” as used herein is defined as the sequence GR, wherein the scissile bond is located immediately after the Arg, and wherein the GR is not immediately followed by a Pro or an Arg. More preferably, the Factor Xa cleavage site is a sequence selected from the group consisting of IDGR (SEQ ID NO:50) and IEGR (SEQ ID NO:51), wherein for each case, the scissile bond is located after the Arg, and wherein the aforementioned sequences are not immediately followed by a Pro or an Arg.

[0029] A “protease domain” as used herein is an amino acid sequence comprising at least 20 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In one embodiment, the protease domain comprises at least 60 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In

another embodiment, the protease domain comprises at least 120 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In another embodiment, the protease domain comprises at least 180 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In yet another embodiment, the protease domain comprises at least 240 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In still another embodiment, the protease domain comprises at least 300 contiguous amino acid from residues 74-446 of SEQ ID NO:1. In another embodiment, the protease domain comprises at least one sequence selected from the group consisting of SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54 and SEQ ID NO:55.

GYVEMTVGSPPQTLNILVDTGSSNFAV	SEQ ID NO:52
SSTYRDLRKGVYVPYTQGWEGELGTDL	SEQ ID NO:53
ESDKFFINGSNWEGILGLAYA	SEQ ID NO:54
EYNYDKSIVDSGTTNLRLP	SEQ ID NO:55

[0030] SEQ ID NO:52 corresponds to residues 74-101 of isoforms A, B, C, and D of human BACE1, and SEQ ID NO:53 corresponds to residues 118-145 of isoforms A, B, C, and D of human BACE1. SEQ ID NO:54 corresponds to residues 165-185 of human BACE1 isoform A and human BACE1 isoform B, and is not present in human BACE1 isoform C or human BACE1 isoform D. SEQ ID NO:55 corresponds to residues 280-298 of human BACE1 isoform A, residues 255-273 of human BACE1 isoform B, residues 236-254 of human BACE1 isoform C, and residues 211-229 of human BACE1 isoform D, respectively. In one embodiment, the protease domain comprises SEQ ID NO:52. In another embodiment, the protease domain comprises SEQ ID NO:53. In another embodiment, the protease domain comprises both SEQ ID NO:52 and SEQ ID NO:53. In another embodiment, the protease domain comprises SEQ ID NO:52, SEQ ID NO:53, or both SEQ ID NO:52 and SEQ ID NO:53, and further comprises at least one sequence selected from the group consisting of SEQ ID NO:54 and SEQ ID NO:55. In yet another embodiment, the protease domain comprises SEQ ID NO:52 and SEQ ID NO:54. In another embodiment the protease domain comprises SEQ ID NO:52 and SEQ ID NO:55. In another embodiment, the protease domain comprises SEQ ID NO:53 and SEQ ID NO:54. In another embodiment, the protease domain comprises SEQ ID NO:53 and SEQ ID NO:55. In another embodiment, the protease domain comprises SEQ ID NO:52, SEQ ID NO:53 and SEQ ID

NO:54. In another embodiment, the protease domain comprises SEQ ID NO:52, SEQ ID NO:53 and SEQ ID NO:55. In another embodiment, the protease domain comprises SEQ ID NO:52, SEQ ID NO:54 and SEQ ID NO:55. In another embodiment, the protease domain comprises SEQ ID NO:53, SEQ ID NO:54, and SEQ ID NO:55. In still another embodiment, the protease domain comprises SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54 and SEQ ID NO:55.

[0031] The term “soluble” as used in reference to the polypeptides herein, is defined as lacking transmembrane and cytoplasmic domains.

[0032] For purposes of shorthand designation of the polypeptides described herein, it is noted that numbers refer to the position of the altered amino acid residue along the amino acid sequences of respective wild-type protein sequence. Amino acid identification uses the single-letter alphabet of amino acids, i.e.

[0033] Asp    D        Aspartic acid

[0034] Ile     I        Isoleucine

[0035] Thr     T        Threonine

[0036] Leu     L        Leucine

[0037] Ser     S        Serine

[0038] Tyr     Y        Tyrosine

[0039] Glu     E        Glutamic acid

[0040] Phe     F        Phenylalanine

[0041] Pro     P        Proline

[0042] His     H        Histidine

[0043] Gly     G        Glycine

[0044] Lys     K        Lysine

[0045] Ala     A        Alanine

[0046] Arg     R        Arginine

[0047] Cys     C        Cysteine

[0048] Trp     W        Tryptophan

[0049] Val     V        Valine

[0050] Gln     Q        Glutamine

[0051] Met     M        Methionine

[0052] Asn     N        Asparagine



[0053] Unless otherwise indicated, numbering of specific amino acids and amino acid sequences herein is with respect to the preprosequence of human BACE1 (SEQ ID NO:1).

[0054] Percent identity of a protein sequence to a reference protein sequence, is determined by alignment of the protein sequence to the reference protein sequence using the GAP program [Huang, X., *Computer Applications in the Biosciences* **10**, 227-235 (1994)] in the Wisconsin Genetics Software Package Release 10.0, with a BLOSUM62 comparison matrix [Henikoff, S. & Henikoff, J. G., *Proc Natl Acad Sci U S A* **89**:10915-10919 (1992)] and default parameters for gap openings and gap extensions. The alignment is calculated over the entire length of the reference sequence, with no penalty for gaps outside the alignment to the reference sequence. The program GAP is an implementation of the Needleman-Wunsch algorithm [Needleman & Wunsch, *J. Mol. Biol.*, **48**, 443-453 (1970)].

**[0055] Embodiments**

[0056] The invention is directed to polypeptides having BACE activity and having at least one engineered cleavage site allowing proteolysis or autoproteolysis of the polypeptide *in vitro*.

[0057] One aspect of the invention is a polypeptide comprising in order from the N-terminus to the C-terminus:

- a) a prodomain comprising at least six contiguous amino acids of SEQ ID NO:3;
- b) an engineered cleavage site; and
- c) a protease domain;

[0058] wherein the polypeptide is capable of being cleaved at the engineered cleavage site thereby releasing a free protease domain that has BACE activity.

[0059] In a particular embodiment of the polypeptide, the prodomain comprises at least seven contiguous amino acids of SEQ ID NO:3. In another embodiment, the prodomain comprises at least eight contiguous amino acids of SEQ ID NO:3. In yet another embodiment, the prodomain comprises at least ten contiguous amino acids of SEQ ID NO:3. In still another embodiment, the prodomain comprises at least twelve contiguous amino acids of SEQ ID NO:3. In a preferred embodiment, the prodomain comprises the full length of SEQ ID NO:3.

[0060] In another embodiment of the polypeptide, the engineered cleavage site is an autoproteolysis site, as defined herein. In another embodiment of the polypeptide, the engineered cleavage site is an exogenous protease cleavage site, as defined herein.

**[0061] Autoproteolysis Sites**

**[0062]** In certain polypeptides of this invention, a polypeptide having BACE activity is engineered for autoproteolysis.

**[0063]** Autoproteolysis sites for BACE1 as used herein generally comprise the amino acid sequence X1-X2-X3-X4-X5, where X1, X2, X3, X4, and X5 correspond to substrate sequence positions P4, P3, P2, P1, and P1', respectively, according to the nomenclature of Schechter and Berger [Schechter, I. & Berger, A., *Biochem. Biophys. Res. Commun.* **27**:157-162 (1967)].

**[0064]** A particular aspect of the invention is a polypeptide comprising in order from the N-terminus to the C-terminus:

- a) a prodomain comprising SEQ ID NO:3;
- b) an autoproteolysis site comprising the sequence X1-X2-X3-X4-X5,

wherein

X1 is Glu or Gln,

X2 is Leu, Ile or Val,

X3 is Asn, Asp or Met,

X4 is Leu or Phe, and

X5 is Glu, Met, Gln, Ser, Ala or Asp; and

- c) a protease domain comprising at least 20 contiguous amino acids from residues 74-446 of SEQ ID NO:1;

wherein the polypeptide is capable of being cleaved at the autoproteolysis site thereby releasing a free protease domain that has BACE activity.

**[0065]** In another embodiment of the polypeptide, the protease domain comprises at least 60 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the protease domain comprises at least 120 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the protease domain comprises at least 180 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the protease domain comprises at least 240 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the protease domain comprises at least 300 contiguous amino acids from residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the protease domain comprises residues 74-446 of SEQ ID NO:1.

[0066] In another embodiment of the polypeptide, the protease domain comprises at least one sequence selected from the group consisting of SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, and SEQ ID NO:55.

[0067] In another embodiment of the polypeptide, the autoproteolysis site is selected from the group

consisting of:

- a) ELNLETD (SEQ ID NO:56),
- b) EINLETD (SEQ ID NO:57),
- c) EINFETD (SEQ ID NO:58), and
- d) EVNLDAE (SEQ ID NO:59).

[0068] In yet another embodiment of the polypeptide, the autoproteolysis site is selected from the group

consisting of:

- a) EINFSEFVE (SEQ ID NO:60),
- b) EINFQFVD (SEQ ID NO:61), and
- c) EINFSAF (SEQ ID NO:62).

[0069] In still another embodiment of the polypeptide, the prodomain comprises residues 22-41 of SEQ ID NO:1. In another embodiment of the polypeptide, the prodomain comprises residues 22-45 of SEQ ID NO:1.

[0070] In another aspect, the invention is directed to a polypeptide comprising an amino acid sequence that is at least 85% identical to residues 22-446 of SEQ ID NO:1 wherein the sequence includes at least one autoproteolysis site X1-X2-X3-X4-X5,

wherein

- X1 is Glu or Gln,
- X2 is Leu, Ile or Val,
- X3 is Asn, Asp or Met,
- X4 is Leu or Phe, and
- X5 is Glu, Met, Gln, Ser, Ala or Asp.

[0071] In one embodiment of the polypeptide the sequence is at least 90% identical to residues 22-446 of SEQ ID NO:1. In another embodiment of the polypeptide the sequence is at least 95% identical to residues 22-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the

sequence is at least 97% identical to residues 22-446 of SEQ ID NO:1. In another embodiment of the polypeptide the sequence is at least 99% identical to residues 22-446 of SEQ ID NO:1.

[0072] In another embodiment of the polypeptide comprising the amino acid sequence that is at least 85% identical to residues 22-446 of SEQ ID NO:1, X1 of the autoproteolysis site is located at least 10 amino acids from the N-terminus of the sequence. In another embodiment, X1 of the autoproteolysis site is located at least 15 amino acids from the N-terminus of the sequence. In another embodiment, X1 of the autoproteolysis site is located at least 20 amino acids from the N-terminus of the sequence.

[0073] In another embodiment of the polypeptide, the amino acid sequence is at least 85% identical to residues 22-446 of SEQ ID NO:1 and includes a subsequence that is at least 90% identical to residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the amino acid sequence is at least 85% identical to residues 22-446 of SEQ ID NO:1 and includes a subsequence that is at least 95% identical to residues 74-446 of SEQ ID NO:1. In another embodiment of the polypeptide, the amino acid sequence is at least 85% identical to residues 22-446 of SEQ ID NO:1 and includes a subsequence that is at least 99% identical to residues 74-446 of SEQ ID NO:1. In still another embodiment of the polypeptide, the amino acid sequence is at least 85% identical to residues 22-446 of SEQ ID NO:1 and includes a subsequence that is at least 99.5% identical to residues 74-446 of SEQ ID NO:1.

[0074] In another aspect, the invention is directed to a polypeptide comprising an amino acid sequence that is at least 85% identical to residues 22-446 of SEQ ID NO:1 wherein the sequence includes

- a) a subsequence that is at least 90% identical to residues 74-446 of SEQ ID NO:1; and
- b) at least one autoproteolysis site X1-X2-X3-X4-X5,

wherein

X1 is Glu or Gln,

X2 is Leu , Ile or Val,

X3 is Asn, Asp or Met,

X4 is Leu or Phe, and

X5 is Glu, Met, Gln, Ser, Ala or Asp; and

wherein X1 is located at least 6 residues from the N-terminus of the amino acid sequence. In another embodiment of the polypeptide, the amino acid sequence is at least 90%

identical to residues 22-446 of SEQ ID NO:1 and the subsequence is at least 95% identical to residues 74-446 of SEQ ID NO:1.

**[0075]** Exogenous protease cleavage

**[0076]** In other polypeptides of the invention, a polypeptide having BACE activity is engineered for exogenous protease cleavage.

**[0077]** An exogenous protease cleavage sites is sequence comprising at least two amino acids that is cleaved by an exogenous protease, a polypeptide having activity other than BACE activity.

**[0078]** In one embodiment, the invention is directed to a polypeptide comprising in order from the N-terminus to the C-terminus:

- a) a prodomain comprising at least six contiguous amino acids of SEQ ID NO:3;
- b) an exogenous protease cleavage site; and
- c) a protease domain;

wherein the polypeptide is capable of being cleaved at the exogenous protease cleavage site thereby releasing a free protease domain that has BACE activity. In a particular embodiment, the exogenous protease cleavage site is a thrombin cleavage site, a tobacco etch virus cleavage site, a Genenase I cleavage site, an Enterokinase cleavage site, a Granzyme B cleavage site, a turnip mosaic virus protease NIa cleavage site, or a Factor Xa cleavage site.

**[0079]** Conditions for the *in vitro* cleavage of the engineered cleavage site by the exogenous protease should be chosen so as to optimize the selectivity of the exogenous protease for its preferred cleavage sites over any other potential or non-preferred sites. Effects of adjustment of certain cleavage conditions are well known in the art. For example, greater specificity is achieved in enzymatic reactions having a low enzyme: substrate ratio. Optimally, the lowest amount of enzyme necessary for complete and specific cleavage should be used. Additionally, enzymes can show greater specificity in substrate cleavage at lower temperatures, for example, at 4 °C relative to at 37 °C.

**[0080]** Engineered Cleavage Sites

**[0081]** An important consideration in designing the engineered cleavage site is the selectivity of the protease to be used with respect to the sequence of the remainder of the polypeptide having BACE activity, especially of its protease domain. For example, in particular embodiments of this invention, exogenous protease cleavage sites of enzymes having well-characterized

selectivity are preferred. Furthermore, it is preferable that the engineered cleavage site is created such that there are few preferred or non-preferred sequences, preferably less than three, that may be cleaved by the exogenous protease or by autoproteolysis occurring downstream of the engineered cleavage site within the polypeptide. It is possible that uniform processing could occur in a polypeptide having BACE activity that contains more than one cleavage site for a given protease, particularly if the engineered cleavage site selected were highly preferred by that protease over other existing sites. Furthermore, it is possible that uniform processing could be effected by a protease where equally preferred sites exist, because only the engineered cleavage site may be accessible to the protease when the polypeptide having BACE activity exists in a folded state. Additionally, it is allowable that additional, non-overlapping sites for the protease to be present upstream (i.e., on the N-terminal side) of the engineered cleavage site in the polypeptide, as long as the engineered cleavage site is more preferred than the additional upstream sites.

[0082] In particular, is preferable to place the engineered cleavage sites within the polypeptide having BACE activity such that there are not overlapping sites spaced closely together, i.e., within less than about 5 amino acids of one another. Cleavage sites that are placed closely enough to overlap may have similar rates of cleavage. For overlapping sites located closer to the N-terminus of the protein, cleavage at the N-terminal cleavage site would prevent cleavage at the C-terminal cleavage site because upon cleavage of the N-terminal site, a portion of the C-terminal site would be removed. Thus overlapping engineered cleavage sites are expected to produce heterogeneous cleavage products.

[0083] In a preferred embodiment of polypeptides having BACE activity and comprising an engineered cleavage site, the polypeptide has no sites occurring within the protease domain that can be cleaved by the corresponding exogenous protease or by autoproteolysis. An exogenous protease cleavage site or autoproteolysis site located within the protease domain is disadvantageous as cleavage at that site may reduce or eliminate the BACE activity of the released product. Exogenous protease cleavage sites or autoproteolysis sites contained within the protease domain may be eliminated through alteration of the protease domain itself so that the site contained therein is removed while BACE activity is retained.

[0084] Specific examples of preferred exogenous protease cleavage sites not occurring in the protease domain of native BACE1 include the TEV protease cleavage site, the Genenase I

cleavage site, the Enterokinase cleavage site, a Granzyme B cleavage site, and the TuMV protease NIa cleavage site. Another example of a preferred engineered cleavage site is a Factor Xa cleavage site. Although it would appear that, for example, BACE1 has a Factor Xa cleavage site at <sup>53</sup>EPGRRG<sup>58</sup> (SEQ ID NO:63), with cleavage occurring after Arg56, the P4 and P3 positions of this site do not conform to a preferred Factor Xa cleavage site. Moreover, it is also known that Factor Xa will not cleave a site followed by Pro or Arg. Examples of more preferred exogenous protease cleavage sites-expected to be targets of highly specific cleavage in the context of BACE1-are preferred TEV protease cleavage sites, preferred Enterokinase cleavage sites, preferred TuMV protease NIa cleavage sites, the preferred Genenase I cleavage site, or preferred Factor Xa cleavage sites.

[0085] It is possible to construct polypeptides having BACE activity and comprising engineered cleavage sites by modification of a native BACE1 sequence. For example, one or more amino acids may be introduced between the N-terminal residue of the prodomain, and the protease domain. In particular, an engineered cleavage site may be composed entirely of a stretch of residues inserted between two residues of this region, for example, see SEQ ID NO:76 (soluble human BACE1 with a preferred thrombin site insert). Alternatively, one or more amino acids may be substituted with different amino acids, e.g., see SEQ ID NO:81. If more than one amino acid is substituted, the amino acids having substitutions may be contiguous or noncontiguous. Another means of introducing the engineered cleavage site involves deletion of one or more amino acids, so that the remaining sequence surrounding the deleted residue or residues constitutes the engineered cleavage site. Finally, any combination of insertion, substitution or deletion of amino acids may be used to introduce the engineered cleavage site at the desired position of BACE1.

**[0086] Cysteine mutants**

[0087] In another aspect, the invention is directed to a polypeptide comprising BACE activity and further comprising a substitution to a cysteine of a native noncysteine residue. Such polypeptides are useful for discovery and development of BACE1 inhibitors by use of Tethering<sup>SM</sup>, which is described in US Patent No. 6,335,155, WO 02/42773, and WO 03/046200. In brief, Tethering<sup>SM</sup> enables discovery of low molecular weight compounds that weakly bind to a region on a protein target, through formation of a covalent bond between the compound and a cysteine residue located at the region. Thus a polypeptide having BACE activity and comprising

an introduced cysteine may be used to screen against a library of compounds each containing a group capable of reacting with the cysteine thiol. Under certain conditions, the subset of compounds having noncovalent binding affinity to the region at which the cysteine is located can form a covalent linkage to the introduced cysteine on the polypeptide. The resulting polypeptide-compound conjugate may subsequently be identified, e.g., by mass spectroscopy. The protease domain of native human BACE1 comprises six cysteines, each of which are involved in disulfide bond formation, as described in the Background. Therefore, it is preferable that polypeptides used to find compounds binding the protease domain by Tethering<sup>SM</sup> comprise an additional, reactive cysteine located in the protease domain. In addition, the polypeptides having BACE activity and comprising a mutation of a noncysteine residue to a cysteine may or may not further comprise an engineered cleavage site.

[0088] Amino acid insertions, substitutions, or deletions used to install a mutation to cysteine or an engineered cleavage site into a polypeptide having BACE activity may be introduced using several techniques. Such techniques include, for example, site-directed mutagenesis of the nucleic acid sequence encoding a polypeptide having BACE activity such that the nucleic acid sequence encodes a polypeptide having BACE activity, and further comprising a mutation to cysteine, an engineered cleavage site, or both. Particularly preferred is site-directed mutagenesis using polymerase chain reaction (PCR) amplification [see, for example, U.S. Pat. No. 4,683,195 issued 28 July 1987; and *Current Protocols In Molecular Biology*, Chapter 15 (Ausubel, F. M., et al., ed., (1991))]. Other site-directed mutagenesis techniques are also well known in the art and are described, for example, in the following publications: Ausubel, F. M., et al., *supra*, Chapter 8; *Molecular Cloning: A Laboratory Manual*, 2nd edition (Sambrook, J., et al., 1989); Zoller, M. J. & Smith, M., *Methods Enzymol.* **100**:468-500 (1983); Zoller, M. J. & Smith, M., *DNA* **3**:479-488 (1984); Zoller, M. J. & Smith, M., *Nucleic Acids Res.*, **10**:6487-6500 (1982); Brake, A. J., et al., *Proc. Natl. Acad. Sci. USA* **81**:4642-4646 (1984); Botstein, D., et al., *Science* **229**:1193-1201(1985); Kunkel, T. A., et al., *Methods Enzymol.* **154**:367-382 (1987), Adelman, J. P., et al., *DNA* **2**:183-193 (1983); and Carter, P., et al., *Nucleic Acids Res.* **13**:4431-43 (1985).

[0089] Amino acid sequence mutants with more than one amino acid substitution located close together in the polypeptide chain may be generated simultaneously, using one oligonucleotide that codes for all of the desired amino acid substitutions. Cassette mutagenesis [Wells, J. A., et al., *Gene*, **34**:315-323 (1985)], and restriction selection mutagenesis [Wells, et al., *Philos. Trans.*



*R. Soc. London SerA*, 317:415 (1986)] may also be used. If, however, the amino acids are located some distance from one another (e.g. separated by more than ten amino acids), it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed. In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. The alternative method involves two or more rounds of mutagenesis to produce the desired mutant.

[0090] Another aspect of the invention is a nucleic acid used to encode the polypeptides having BACE activity and having an engineered cleavage site. The design of these nucleic acids is influenced by the choice of host cell, as mentioned above. It is well known in the art that as a consequence of the degenerate genetic code, different organisms can have distinct preferences of codon usage. Nucleic acids of the invention encoding the polypeptide are preferably optimized for translation of the encoded polypeptide in the particular host chosen, using information known in the art about codon preferences for the host organism. The codon usage preferences of hundreds of species have been catalogued [see [www.kazusa.or.jp/codon/](http://www.kazusa.or.jp/codon/); Nakamura, Y., et al., *Nucl. Acids Res.* 28: 292 (2000)]. For example, *E. coli* codon usage preferences have been particularly well characterized [Ikemura, T., *J. Mol. Biol.* 151:389-409 (1981); Blake, R. D. & Hinds, P. W., *J. Biomol. Struct. Dynam.* 2:593-606 (1984); Hernan, R. A., et al., *Biochemistry* 31:8619-8627 (1992)]. Optimal polypeptide expression in *E. coli*, can be accomplished by the use of silent codon substitution mutagenesis to replace codons used more frequently for expression in human cells with codons used more frequently for expression in *E. coli*. One embodiment of the invention is a nucleic acid encoding a polypeptide comprising:

- a) a prodomain comprising at least six contiguous amino acids of SEQ ID NO:3;
- b) an engineered cleavage site; and
- c) a protease domain;

wherein the polypeptide is capable of being cleaved at the engineered cleavage site thereby releasing a free protease domain that has BACE activity.

[0091] Another aspect of the invention is a vector for expressing the polypeptides having BACE activity and having an engineered cleavage site. These vectors comprise a nucleic acid encoding the polypeptide, operably linked to an expression control sequence. Expression and cloning

vectors are well known in the art and contain nucleic acid sequences that enable the vector to replicate in one or more selected host cells. Expression vectors including the expression control sequences, e.g., promoters, to be used in the creation of the constructs are preferably optimized for use in the particular host cell chosen. There are numerous commercially available expression vectors for production of protein in a variety of cell types. The vector components generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence. Furthermore, Shine-Dalgarno consensus sequences may be employed preceding the start codon to increase translation efficiency of a eukaryotic gene product in a prokaryotic host cell. A particular embodiment is a vector for expressing a polypeptide comprising:

- a) a prodomain comprising at least six contiguous amino acids of SEQ ID NO:3;
- b) an engineered cleavage site; and
- c) a protease domain;

wherein the polypeptide is capable of being cleaved at the engineered cleavage site thereby releasing a free protease domain that has BACE activity. A suitable vector is described in Example 1.

[0092] Another aspect of the invention is a host cell expressing a polypeptide having BACE activity and having an engineered cleavage site. Host cells are transformed with the expression or cloning vector encoding the polypeptide, and are cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. In one embodiment, the host cell expresses a polypeptide having BACE activity and comprising an autoproteolysis site. In another embodiment the host cell expresses a polypeptide having BACE activity and comprising an exogenous protease cleavage site. Preferably, the polypeptide comprising an exogenous protease cleavage site is expressed in host cells not natively expressing the corresponding exogenous protease, because isolating the polypeptide comprising an intact exogenous protease cleavage site allows processing to occur under more controlled, *in vitro* conditions. If the polypeptide having BACE activity is to be used for structural studies, it preferably is expressed in prokaryotic cells, more preferably in bacterial cells, and most preferably, in *E. coli* cells. Expressing the polypeptides in these cell types avoids the heterogeneous glycosylation observed in polypeptides isolated from

eukaryotic systems. In a particular embodiment, the host cell expresses a polypeptide comprising:

- a) a prodomain comprising at least six contiguous amino acids of SEQ ID NO:3;
- b) an engineered cleavage site; and
- c) a protease domain;

wherein the polypeptide is capable of being cleaved at the engineered cleavage site thereby releasing a free protease domain that has BACE activity. Example 2 describes the purification from *E. coli* of polypeptides having BACE activity and having an autoproteolysis site.

#### EXAMPLES

##### [0093] EXAMPLE 1

##### [0094] CONSTRUCTION OF BACTERIAL EXPRESSION PLASMIDS ENCODING POLYPEPTIDES WITH BACE ACTIVITY AND WITH ENGINEERED CLEAVAGE SITES

[0095] In brief, a pRSETC (Novagen)-based *E. coli* expression plasmid, pB22, encoding human proBACE1 starting at amino acid 22 from the wild-type N-terminus, was used as a template for introduction of mutations by Kunkel mutagenesis [Kunkel, T. A., et al., *Methods Enzymol.* **154**:367-382 (1987)]. The sequence encoding the prodomain and linker had been previously optimized for *E. coli* expression by silent codon substitution mutagenesis.

##### [0096] *Cloning of Soluble Human BACE1*

[0097] A soluble proprotease gene sequence (bases 64-1362, amino acid residues 22-454 of SEQ ID NO:1) was subcloned from pFBHT into the *E. coli* expression vector pRSETC by PCR, to create pB22, which served as a template for mutagenesis. pFBHT is a modified pFastBac1 plasmid (Gibco/BRL) containing the sequence for TEV protease followed by a (His)<sub>6</sub> tag and a stop signal between the XhoI and HindIII sites. The subcloning was accomplished as follows. The cDNA encoding full-length human BACE1, bases 1-1551, starting from the initiator Met codon and including an extra 48 bases of mRNA transcript following the stop codon [Vassar, R., et al., *Science* **286**:735-741 (1999)] was obtained by a combination of PCR cloning of the 3' 1425 bases from human cDNA libraries, and synthesis of the remaining 5' 126 bases by serial overlapping PCR. All PCR reactions were performed using Advantage2 polymerase (Clontech) according to manufacturer's instructions. A fragment spanning bases 126-374 was obtained by PCR from a human cerebral cortex library and SEQ ID NO:64 and SEQ ID NO:65; a fragment spanning bases 339-770 was obtained by PCR from a Stratagene Unizap XR human brain cDNA

library, and SEQ ID NO:66 and SEQ ID NO:67; and the 3' end fragment, spanning bases 735-1551, was obtained by PCR from a human brain library, using SEQ ID NO:68 and SEQ ID NO:69. The three fragments, having 35 bp of overlap at the junctions, were gel purified and combined in one PCR reaction, using primers to the ends (SEQ ID NO:64 and SEQ ID NO:70) to amplify the 126-1551 product.

For2	GCTGCCCCGGGAGACCGACGAAGA	SEQ ID NO:64
midRev2	CGGAGGTCCCGGTATGTGCTGGAC	SEQ ID NO:65
midFor	CCAGAGGCAGCTGTCCAGCACATA	SEQ ID NO:66
midRev1	TCCCGCCGGATGGGTGTATACCAG	SEQ ID NO:67
BACE14	GTACACAGGCAGTCTCTGGTATACACC	SEQ ID NO:68
BACE11	GTGTGGTCCAGGGGAATCTCTATCTTCTG	SEQ ID NO:69
BACE5	GTCATCGTCTCGAGTCACTTCAGCAGGGAGATGTCATCAG	SEQ ID NO:70

[0098] The 126-1551 piece, and the subsequent elongated products, were used as a templates for serial overlapping PCR reactions, to add the remaining 5' –126 bases using SEQ ID NO:71, SEQ ID NO:72 and SEQ ID NO:73 as forward primers, with SEQ ID NO:75 always at the reverse primer.

[0099] BACE fill2

CGGCTGCCCCTGCGCAGCGGCCTGGGGGGCGCCCCCTGGGGCTGCGGCTGCCCCGGGAG  
SEQ ID NO:71

[0100] BACE fill1

ATGGGCGCGGGAGTGCTGCCTGCCCACGGCACCCAGCACGGCATCCGGCTGCCCCTGCGC  
SEQ ID NO:72

[0101] BACE for-EcoRI

CCGGAATTCATGGCCCAAGCCCTGCCCTGGCTCCTGCTGTGGATGGGCGCGGGAGTG  
SEQ ID NO:73

[0102] SEQ ID NO:73 and SEQ ID NO:70 contained EcoRI and XhoI restriction sites, respectively, and digestion of the PCR product, along with the Baculovirus expression vector,

pFBHT, with the same enzymes was followed by gel purification and ligation of the resulting DNA fragments, yielding the construct, pFBHT-BACE. This construct was used as a template for PCR amplification of bases 1-1362, corresponding to the preproBACE soluble protease, using SEQ ID NO:74 and SEQ ID NO:75.

proFor-Nde CGCCATATGGCGGGAGTGCTGCCTGCCCACGGC SEQ ID NO:74

BACErev-RI CCGGAATTCTCAGGTTGACTCATCTGTCTGTGGAAT SEQ ID NO:75

[0103] SEQ ID NO:74 and SEQ ID NO:75 contained NdeI and EcoRI restriction sites, respectively, and digestion of the PCR product, along with the *E. coli* expression vector, pRSETC, with the same enzymes was followed by gel purification and ligation of the resulting DNA fragments leading to the construct pB1. Vector pB1 was then used as a template for Kunkel mutagenesis [Kunkel, T. A., et al., *Methods Enzymol.* **154**:367-382 (1987)] to delete the BACE1 presequence (bases 1-63), producing the construct pB22. pB22 served as a template for mutagenesis to incorporate the engineered cleavage sites, using the Kunkel method.

[0104] *Introduction of engineered cleavage sites*

[0105] Mutations and the oligonucleotides used to introduce them are as follows. For the construct encoding a polypeptide of SEQ ID NO:76, the thrombin cleavage site LVPRGS (SEQ ID NO:4) was inserted between residues 45 and 46 of the aforementioned soluble proBACE1, numbered according to the preprosequence in SEQ ID NO:1. The residues correspond to residues 24 and 31, respectively, of SEQ ID NO:76. The oligonucleotide used for this insertion was SEQ ID NO:77.

TQHGIRLPLR SGLGGAPLGL RLPRLVPRGS ETDEEPEEPG RRGSFVEMVD SEQ ID NO:76  
 NLRGKSGQGY YVEMTVGSPP QTLNILVDTG SSNFAVGAAP HPFLHRYYQR  
 QLSSTYRDLR KGVYVPYTQG KWEGELGTDL VSIPHGPNVT VRANIAAITE  
 SDKFFINGSN WEGILGLAYA EIARPDDSLE PFFDSL VKQT HVPNLFSLQL  
 CGAGFPLNQS EVLASVGGSIM IIGGIDHSLY TGSLWYTPIR REWYVEVIIV  
 RVEINGQDLK MDCKEYNKYDK SIVDSGTTNL RLPKKVFEAA VKSIKAASST  
 EKFPDGFVLG EQLVCWQAGT TPWNIFPVIS LYLMGEVTNQ SFRITILPQQ  
 YLRPVEDVAT SQDDCYKFAI SQSSTGTVMG AVIMEGFYVV FDRARKRIGF  
 AVSACHVHDE FRTAAVEGPF VTLDMEDCGY NIPQTDEST

thrombin CTCTTCGTCGGTCTCAGAACACGCGGAACCAGACGTGGCAGACGCAG SEQ ID NO:77  
 site insert

[0106] Another construct encodes for a polypeptide of SEQ ID NO:78 containing the sequence RLPLETD (SEQ ID NO:79). The sequence was created by a single point mutation to a Leu of residue Arg45 of proBACE1, numbered relative to the preprosequence. The mutated residue corresponds to residue 24 of SEQ ID NO:78. The oligonucleotide used to make the amino acid substitution was SEQ ID NO:80.

TQHGIRLPLR	SGLGGAPLGL	RLPLETDEEP	EEPGRRGSFV	EMVDNLRGKS	SEQ ID NO:78
GQGYVEMTV	GSPPQTLNIL	VDTGSSNFAV	GAAPHPFLHR	YYQRQLSSTY	
RDLRKGVYVP	YTQGWEGEL	GTDLVSIHPG	PNVTVRANIA	AITESDKFFI	
NGSNWEGILG	LAYAEIARPD	DSLEPFFDSL	VKQTHVPLNF	SLQLCGAGFP	
LNQSEVLASV	GGSMIIGGID	HSLYTGSLWY	TPIRREWYVE	VIIIVRVEING	
QDLKMDCKEY	NYDKSIVDSG	TTNLRPLPKV	FEAAVKSIA	ASSTEKFPDG	
FWLGEQLVCW	QAGTTPWNIF	PVISLYLMGE	VTNQSFRITI	LPQQYLRPVE	
DVATSQDDCY	KFAISQSSTG	TVMGAVIMEG	FYVVFDRARK	RIGFAVSACH	
VHDEFRTAAV	EGPFVTLDE	DCGYNIPQTD	EST		

RLPL site	CTCTTCGTCGGTCTCCAGTGGCAGACGCAGACCCAGTGGAGC	SEQ ID NO:80
-----------	--	--------------

[0107] Another construct encodes for a polypeptide of SEQ ID NO:81 containing the autoproteolysis site ELNLETD (SEQ ID NO:56). The autoproteolysis site was produced by point mutations of three amino acid residues of proBACE1 –Arg42, Pro44, and Arg45-numbered relative to the preprosequence. The residues were mutated to a Glu, an Asp, and a Leu, respectively, which correspond to residues 21, 23, and 24 of SEQ ID NO:81. The oligonucleotide used to make the three point mutations was SEQ ID NO:82.

TQHGIRLPLR	SGLGGAPLGL	ELNLETDEEP	EEPGRRGSFV	EMVDNLRGKS	SEQ ID NO:81
GQGYVEMTV	GSPPQTLNIL	VDTGSSNFAV	GAAPHPFLHR	YYQRQLSSTY	
RDLRKGVYVP	YTQGWEGEL	GTDLVSIHPG	PNVTVRANIA	AITESDKFFI	
NGSNWEGILG	LAYAEIARPD	DSLEPFFDSL	VKQTHVPLNF	SLQLCGAGFP	
LNQSEVLASV	GGSMIIGGID	HSLYTGSLWY	TPIRREWYVE	VIIIVRVEING	
QDLKMDCKEY	NYDKSIVDSG	TTNLRPLPKV	FEAAVKSIA	ASSTEKFPDG	
FWLGEQLVCW	QAGTTPWNIF	PVISLYLMGE	VTNQSFRITI	LPQQYLRPVE	
DVATSQDDCY	KFAISQSSTG	TVMGAVIMEG	FYVVFDRARK	RIGFAVSACH	
VHDEFRTAAV	EGPFVTLDE	DCGYNIPQTD	EST		

ELNL site	CTCTTCGTCGGTCTCCAGGTTCA GTTCCAGACCCAGTGGAGC	SEQ ID NO:82
-----------	---	--------------

[0108] A construct encoding a polypeptide with the autoproteolysis site EINLET D (SEQ ID NO:57) was created from the ELNL-BACE1 construct by using the oligonucleotide of SEQ ID NO:83, which replaces the first Leu of the autoproteolysis site of SEQ ID NO:56 with an Ile. The resulting construct encoded for a polypeptide of SEQ ID NO:84.

EINL site GGTCTCCAGGTTGATTTCCAGACCCAG

SEQ ID NO:83

TQHGI R L P L R S G L G G A P L G L E I N L E T D E E P E E P G R R G S F V E M V D N L R G K S  
G Q G Y Y V E M T V G S P P Q T L N I L V D T G S S N F A V G A A P H P F L H R Y Y Q R Q L S S T Y  
R D L R K G V Y V P Y T Q G K W E G E L G T D L V S I P H G P N V T V R A N I A A I T E S D K F F I  
N G S N W E G I L G L A Y A E I A R P D D S L E P F F D S L V K Q T H V P N L F S L Q L C G A G F P  
L N Q S E V L A S V G G S M I I G G I D H S L Y T G S L W Y T P I R R E W Y Y E V I I V R V E I N G  
Q D L K M D C K E Y N Y D K S I V D S G T T N L R L P K K V F E A A V K S I K A A S S T E K F P D G  
F W L G E Q L V C W Q A G T T P W N I F P V I S L Y L M G E V T N Q S F R I T I L P Q Q Y L R P V E  
D V A T S Q D D C Y K F A I S Q S S T G T V M G A V I M E G F Y V V F D R A R K R I G F A V S A C H  
V H D E F R T A A V E G P F V T L D M E D C G Y N I P Q T D E S T

SEQ ID NO:84

## [0109] EXAMPLE 2

### [0110] BACE PROTEIN PREPARATION USING ELNL-BACE1 AND EINL-BACE1 CONSTRUCTS

[0111] The expression plasmid encoding the polypeptide ELNL-BACE1 (SEQ ID NO:81) or EINL-BACE1 (SEQ ID NO:84) was transformed into BL21star (DE3) pLysS *E. coli* (Invitrogen) and plated onto an LB/amp plate containing 100 µg/mL ampicillin. A single colony from the plate was used to inoculate a 5 mL culture of 2x YT broth containing 50 µg/mL carbenicillin. The culture was grown at 37 °C until the OD<sub>600</sub> reached between 0.4. and 0.6. Glycerol was added to 15% by volume and the resulting glycerol stocks were frozen in 500 µL aliquots in liquid nitrogen. Aliquots were stored at -80 °C. Large-scale expression cultures were started by scraping the glycerol stock into 400 mL 2YT + 50 µg/mL carbenicillin. Following overnight growth at 37 °C, 50 mL of the culture was used to inoculate 1.5 L of the same media, and after growth at 37 °C to an OD at 600 nm of between 0.7 and 0.8, IPTG was added to a final concentration of 1.0 mM and the induced cultures were grown at 4 h at 37 °C.

[0112] Cells were harvested by centrifugation at 4K rpm. Cell pellets were resuspended in 100 mL buffer TE (10 mM Tris-HCl, 1 mM EDTA pH 8.0) and lysed using a microfluidizer. The crude extract, containing the protein as insoluble inclusion bodies, was centrifuged at 10K rpm

for 10 min, and the resulting pellet washed by resuspension in PBST (1X phosphate buffered saline pH 7.4 [10 mM sodium phosphate, 150 mM NaCl] with 0.5% TritonX-100) followed by centrifugation at 10K rpm for 10 min.

[0113] Washed inclusion body pellets were solubilized in a urea solution (50 mM CAPS pH 10, 8 M urea, 1 mM EDTA, and 100 mM  $\beta$ -mercaptoethanol) with light agitation or rocking for at least 30 min at room temperature, and remaining insoluble debris was removed by centrifugation in a SS-34 rotor at 18K rpm for 45 min. The supernatant, which contains denatured IBs, was removed and its absorbance at 280 nm was measured by spectrophotometry. The supernatant was diluted by addition of the urea solution until the OD<sub>280</sub> absorbance measured between 10 and 12.

[0114] Urea-solubilized ELNL-BACE1 or EINL-BACE1 was refolded by rapid dilution into 50 volumes of rapidly stirred 5 mM Na<sub>2</sub>CO<sub>3</sub>, pH 10, followed by incubation at room temperature for 3-5 days. When BACE1 enzymatic activity no longer increased over time, the protein was concentrated 10-20 fold using an ultrafiltration/diafiltration system (Cole-Parmer) and buffer exchanged or dialyzed into 4 mM Tris, pH 8.0, 50 mM NaCl and loaded onto a Q-Sepharose column. Protein was eluted using a linear gradient of 0 to 0.5 M NaCl in 4 mM Tris-HCl pH 8.0 over 8 column volumes. At this point, for the ELNL-BACE1 construct, mass spectral analysis indicated processing to yield an N-terminus starting at residue 42, numbered relative to the preprosequence (corresponding to residue 21 of SEQ ID NO:81). ELNL-BACE1 and EINL-BACE1 were further purified by S-Sepharose chromatography at pH 4.5. Mass spectral analysis for each of the constructs after the S-Sepharose chromatography indicated processing had occurred to yield an N-terminus starting at residue 46, numbered relative to the preprosequence (corresponding to residue 25 of SEQ ID NO:81 and residue 25 of SEQ ID NO:84). The mass spectrum for the processed product of EINL-BACE1 is shown in Figure 1.

[0115] Subsequently, the purified enzyme was dialyzed versus 20 mM Tris pH 7.5 at 4 °C. Protein concentrations were determined by absorbance at 280 nm, using  $\epsilon_{280}^{1\%} = (0.74)$ . Extinction coefficients were calculated using the Vector NTI software (Informax, Invitrogen Inc.). If protein is to be frozen, glycerol should be added to 15% by volume.



[0116] EXAMPLE 3

[0117] CONSTRUCTION OF BACTERIAL EXPRESSION PLASMIDS ENCODING

[0118] BACE1 POLYPEPTIDES WITH MUTATIONS TO CYSTEINE

[0119] The following mutations were introduced by Kunkel mutagenesis into the ELNL-BACE1 construct, the cloning of which is described in Example 1. All residues are numbered relative to preprosequence of human BACE1 (SEQ ID NO:1).

[0120] Oligonucleotides for BACE Cysteine Mutations

L91C	GCCTGTATCCACACAGATGTTGAGCGT	SEQ ID NO:85
D93C	ACTGCTGCCTGTACACACCAGGATGTT	SEQ ID NO:86
T133C	CCACTTGCCCTGACAGTAGGGCACATA	SEQ ID NO:87
Q134C	TTCCCACTTGCCACAGGTGTAGGGCAC	SEQ ID NO:88
K168C	GTTGATGAAGAAACAGTCTGATTCAGT	SEQ ID NO:89
F169C	GCCGTTGATGAAACACTTGTCTGATTC	SEQ ID NO:90
I171C	GTTGGAGCCGTTACAGAAGAACTTGTC	SEQ ID NO:91
W176C	CAGGATGCCTTCACAGTTGGAGCCGTT	SEQ ID NO:92
I179C	GGCCAGCCCCAGACAGCCTTCCCAGTT	SEQ ID NO:93
I187C	GTCAGGCCTGGCACACTCAGCATAGGC	SEQ ID NO:94
R189C	GGAGTCGTCAGGACAGGCAATCTCAGC	SEQ ID NO:95
Y259C	GATCACCTCATAACACCACTCCCGCCG	SEQ ID NO:96
K285C	GTCCACAATGCTACAGTCATAGTTGTA	SEQ ID NO:97
I287C	GCCACTGTCCACACAGCTCTTGTCATA	SEQ ID NO:98
D289C	GGTGGTGCCACTACACACAATGCTCTT	SEQ ID NO:99
T292C	ACGAAGGTTGGTACAGCCACTGTCCAC	SEQ ID NO:100
N294C	GGGCAAACGAAGACAGGTGGTGCCACT	SEQ ID NO:101
R296C	TTTCTTGGGCAAACAAAGGTTGGTGGT	SEQ ID NO:102
G325C	CACCAGCTGCTCCACTAGCCAGAAACC	SEQ ID NO:103
R368C	ATCTTCCACTGGACACAGGTATTGCTG	SEQ ID NO:104
K382C	TGAGATGGCAAACAGTAACAGTCGTC	SEQ ID NO:105
S386C	CGTGGATGACTGACAGATGGCAAACCTT	SEQ ID NO:106
T390C	CATAACAGTGCCACAGGATGACTGTGA	SEQ ID NO:107
V393C	CAGCTCCCATACAAGTGCCCGTGGATG	SEQ ID NO:108

[0121] Also made in the ELNL-BACE1 background were inactive polypeptides comprising catalytic site mutations.

**[0122] Oligonucleotides for Catalytic Site Knock-outs**

D90N	ACTGCTGCCTGTGTTCAACAGGATGTT	SEQ ID NO:109
D289N	GGTGGTGCCACTGTTCACAATGCTCTT	SEQ ID NO:110

**[0123] EXAMPLE 4**

**[0124] BACE1 CYSTEINE MUTANT PROTEIN PREPARATION FROM CONSTRUCTS  
WITH ENGINEERED CLEAVAGE SITES**

**[0125] Expression:**

[0126] Plasmids expressing human BACE1 having both an autoproteolysis site and a mutation of a native noncysteine residue to cysteine, were separately transformed into BL21 Star (DE3) pLysS and plated on an LB/amp plate (100 µg/mL ampicillin). The plates were grown overnight, after which a freshly transformed colony was used to inoculate a 5 mL culture of 2YT + 50 µg/mL carbenicillin. The culture was grown at 37 °C until the OD<sub>600</sub> reached between 0.4 and 0.6. Glycerol was added to 15% by volume and the resulting glycerol stocks were frozen in 500 µL aliquots in liquid nitrogen. Aliquots were stored at -80 °C.

[0127] Large-scale expression cultures were started by scraping the glycerol stock into 400 mL 2YT + 50 µg/mL carbenicillin, and growing overnight at 37 °C. A 50 mL portion of the overnight culture was used to start at least one 1.5 L culture in a shaker flask, which was grown to an OD<sub>600</sub> near 0.7 or 0.8. At this point, the culture was induced by adding IPTG to a final concentration of 1 mM. The induced cultures were next grown 4 hr at 37 °C, although growing longer is also acceptable. Cells were harvested by centrifugation at 6K rpm.

**[0128] Refolding/Purification:**

**[0129] Inclusion Body Prep**

[0130] Cell pellets were resuspended in a minimal amount of TE pH 8.0 (10 mM Tris-HCl, 1 mM EDTA pH 8.0), and lysed fully using either a sonicator or a microfluidizer. A sample of the whole cell lysate was saved to run on an analytical gel. The whole cell lysate was centrifuged 10 min at 10K rpm, after which the supernatant was removed. A sample of the soluble TE supernatant was also saved for the gel.

[0131] The pellet was resuspended in 200 mL PBST (1X phosphate buffered saline pH 7.4 [10 mM sodium phosphate, 150 mM NaCl] with 0.5% TritonX-100). The resuspension was centrifuged for 10 min at 10K rpm, and washed again with the PBST. A sample of the PBST wash was saved for the gel.

[0132] The inclusion body (IB) pellet was resuspended in a volume of a urea solution (8 M Urea, 50 mM CAPS pH 10, 100 mM BME, 1 mM EDTA) such that the resuspension had a final OD<sub>280</sub> of 8-12 after the subsequent steps of denaturing the protein in the urea and spinning out the insoluble materials. The volume of resuspension therefore depends on the amount of protein present in the IBs. It is convenient to start with less urea solution, and add more if necessary. The pellet in urea solution was incubated with light agitation or rocking for at least 30 min at room temperature, and then centrifuged in SS-34 rotor, for 45 min at 18K rpm. The supernatant, which contains denatured IBs, was saved. At this point, the OD was measured, and, if necessary, urea was added to bring the value to the desired OD<sub>280</sub> range of 8-12. During the denaturation of the protein in the urea solution, the introduced cysteine typically forms an adduct with the BME present in the urea solution. The protein-BME adduct in the supernatant can be frozen at -20 °C for later use.

#### **[0133] Refolding**

[0134] A fresh solution of 0.25 M sodium carbonate pH 10 was used to make up a refolding buffer containing 5 mM Na<sub>2</sub>CO<sub>3</sub> pH 10. The protein in the urea solution having an OD<sub>280</sub> of 8-12 was diluted 1:50 by volume into the refolding buffer by quickly adding the urea solution to the larger volume of 5 mM Na<sub>2</sub>CO<sub>3</sub> pH 10, as the larger volume was stirred on a stir-plate. The activity of the BACE protein, i.e. the BACE-BME adduct, was monitored over the next 3-5 d. We have noted the pH of the solution drops to approximately 9.8 after protein addition, and ends at 9.5 over the course of 5 d. The protein in the refolding buffer was concentrated, and its buffer exchanged using an ultrafiltration/diafiltration system (Masterflex L/S, Cole-Parmer). Generally, concentrating 10-20 fold gave excellent results with little to no precipitation in the solution. Exchange of buffer can be accomplished by diafiltration against 2-3 L Q load buffer (4 mM Tris pH 8.0, 50 mM NaCl) or if more convenient, dialyzing against the Q load buffer.

#### **[0135] Q sepharose FF**

[0136] The protein sample was prepared in the same buffer used to equilibrate and wash the column, that is it was dialyzed into buffer A (4 mM Tris pH 8.0, 50 mM NaCl). For example, if

the protein was present in 2 L of refold solution, it was dialyzed twice against 20 L over 7 or 8 h. The protein solution was removed from the dialysis bag and filtered before loading the column.

[0137] Before loading, the column was washed first in high salt (1-2 M) until the UV absorbance leveled out and then in the protein loading buffer (Buffer A) until the UV absorbance leveled out. The high salt buffer was usually Buffer B (4 mM Tris pH 8.0, 1 M NaCl), which was subsequently used to elute the protein from the column. This column preparation step can be done before every run or after every run, but is preferably done after the run.

[0138] While loading the protein onto the column, the flow rate should not exceed 5 mL/min; with large volumes it is helpful to set up a slow overnight load. As the sample was drawn up into the tubing, it was watched, and stopped just before air was drawn into the tube. The flow through from the waste line is preferably collected during loading, as sometimes the protein does not stick to the column and can be found in the flow through. After the protein was loaded, the column was washed with 4-5 column volumes of Buffer A. The UV absorbance on the chromatogram should level out before starting the elution gradient.

[0139] Although there are several types of gradients that may be employed, for BACE cysteine mutant purification on a 20 mL Q seph FF column, the following linear gradient was typically used: 0-75% Buffer B, 4 mL/min over 40 min, collected in 7 mL fractions, and followed by a short 100% Buffer B wash at the end of the gradient. The fractions were collected during the wash step as well. The fractions were checked for refolded protein by running a non-reducing 10% Bis-Tris PAGE with MOPS buffer, and the appropriate fractions were combined and dialyzed overnight into dH<sub>2</sub>O.

#### **[0140] Acidification**

[0141] The protein was next buffered by adding 1 M sodium acetate pH 4.5, to a final concentration of 20 mM. In general, the addition is best done quickly while mixing the solution. The resulting solution was left for 5-10 min with stirring and any resulting precipitation was filtered. The pH drop can first be done on a small scale (0.5-1 mL). After ensuring that the appropriate material fell out of solution by gel analysis of the supernatant, the remainder of the preparation was acidified. The acidified protein was filtered first using a 0.45 µm filter, and then using a 0.22 µm filter.

**[0142] S Sepharose column**

[0143] At this point, the protein was characterized by assaying activity, by mass spectrometry and by polyacrylamide gel electrophoresis (PAGE). If purity was satisfactory at this stage, the protein was dialyzed into reducing buffer, as described below. If the purity was unsatisfactory, the protein was further purified on a 5 mL S sepharose fast-flow column. Loading flow rate depends on protein concentration; ~1 mg protein/min was used in general. The protein was loaded in 10 mM sodium acetate pH 4.5, and next a linear gradient from 0-100% at 2 mL/min, over 20 min was run, with collection of 3 mL fractions. Fractions containing protein were pooled and dialyzed into reducing buffer in a timely manner, because active BACE autoproteolyzes. Before dialyzing, the OD<sub>280</sub> of the protein was measured, using 0.5 M NaOAc as a blank.

**[0144] Reduction of Cysteine Mutant Proteins**

[0145] The purified protein was dialyzed into reducing buffer containing 20 mM Tris pH 7.5, 125 mM NaCl, 1 mM DTT. The protein was checked periodically by mass spectrometry to monitor removal of  $\beta$ -mercaptoethanol (BME) from the cysteine-BME adduct formed during the interaction with the urea solution containing the BME; the removal was effected by a reduction reaction with the DTT. If the BME was slow to reduce, the dialysis buffer DTT concentration was increased to 5 mM. If the BME was particularly slow to reduce, one of the following two procedures was followed. The first procedure was to add 0.4 M urea to the reducing buffer and to continue checking reduction status. The second procedure was to exchange the protein into 20 mM Tris pH 7.5 thereby removing all of the DTT and urea, and then to add TCEP to a final concentration of 2 mM, by direct addition from a 0.5 M TCEP stock. In either case, the reduction status was monitored. If using TCEP or 5 mM DTT, the number of cysteine modifications by cystamine or by a suitably reactive compound was measured to ensure that the native disulfides were not reduced as well. After the protein was reduced, it was dialyzed into storage buffer (20 mM Tris pH 7.5). If protein was to be frozen, glycerol was added to 15% by volume. Protein concentrations were determined by absorbance at 280 nm, using an  $\epsilon_{280}$  as calculated using Vector NTI software (Informax, Invitrogen Inc.); typical  $\epsilon_{280}$ s for cysteine mutants range from 0.70 to 0.72.